教師なし線形次元削減（13章）

杉山将・本多淳也
sugi@k.u-tokyo.ac.jp, jhonda@k.u-tokyo.ac.jp
http://www.ms.k.u-tokyo.ac.jp

次元の呪い

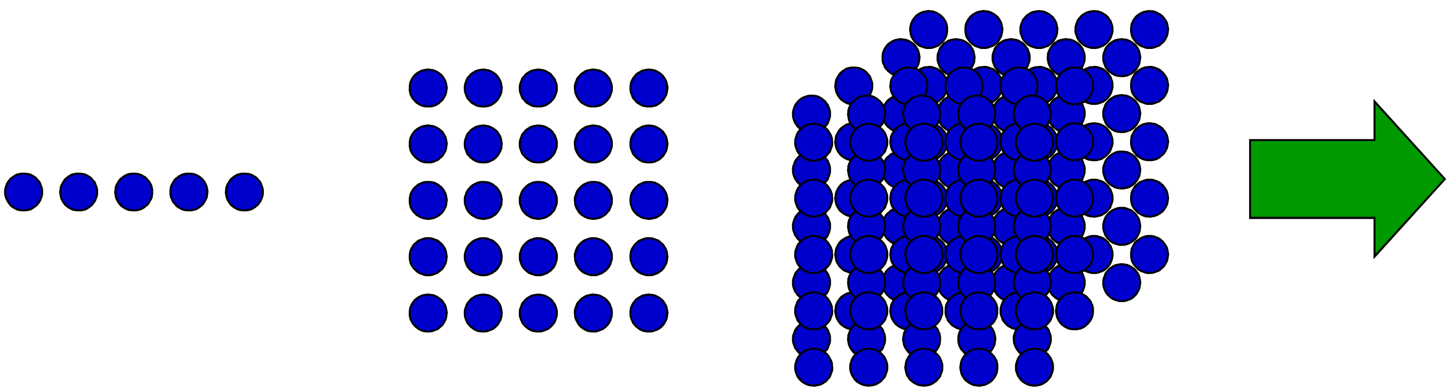
入力の次元dが増えると
学習問題は指数的に難しくなる

- 高次元空間では，通常の幾何学的な直感が成り立たないことがある．

等間隔標本化

- 等間隔標本化に必要な標本数は，
次元数と共に指数的に増加

$d=1$ $d=2$ $d=3$ $\cdots$ d



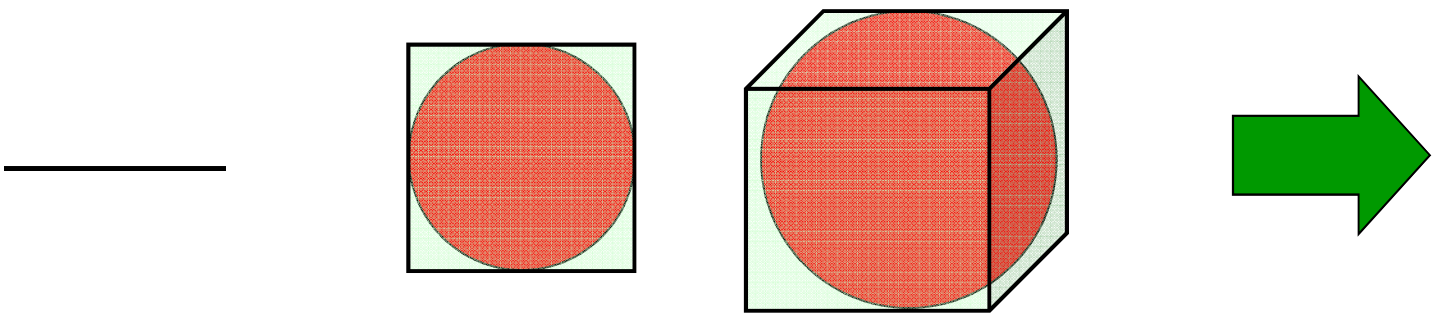
$n=5$ $n=5^2$ $n=5^3$ $n=5^d$

高次元空間では
標本はいつもスパース

内接球の体積

- 次元数が増えていくとき,
  - 単位立方体の体積は常に1
  - 内接球の体積は0に収束

$d=1$ $d=2$ $d=3$ $\cdots$ d



$V=1$

$V=\pi(0.5)^2 \approx 0.79$

$V=4\pi(0.5)^3/3 \approx 0.52$

$V \ll 1$

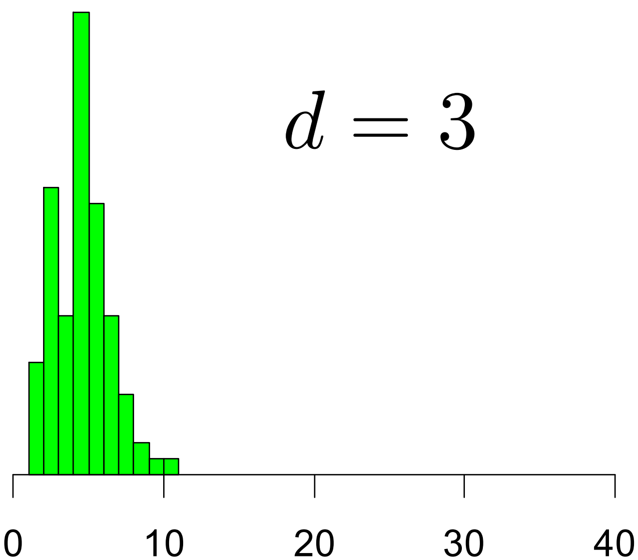
内接球の体積は指数的に小さくなっていく！

$$V \propto (0.5)^d$$

例:k近傍グラフの次数

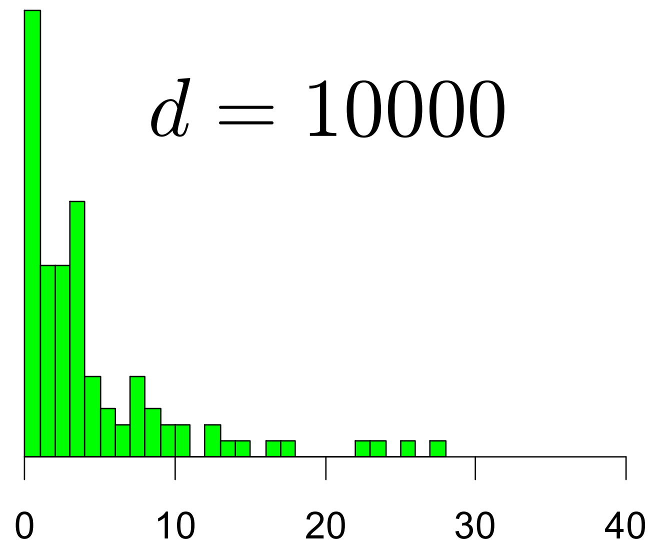
- 超立方体上に n 個の点を一様分布にしたがってランダムに配置
- 各点がそれぞれ何個の点からの k 近傍になっているかをカウント $(n=100, k=5)$

$d=3$



近傍数は k 程度

$d=10000$

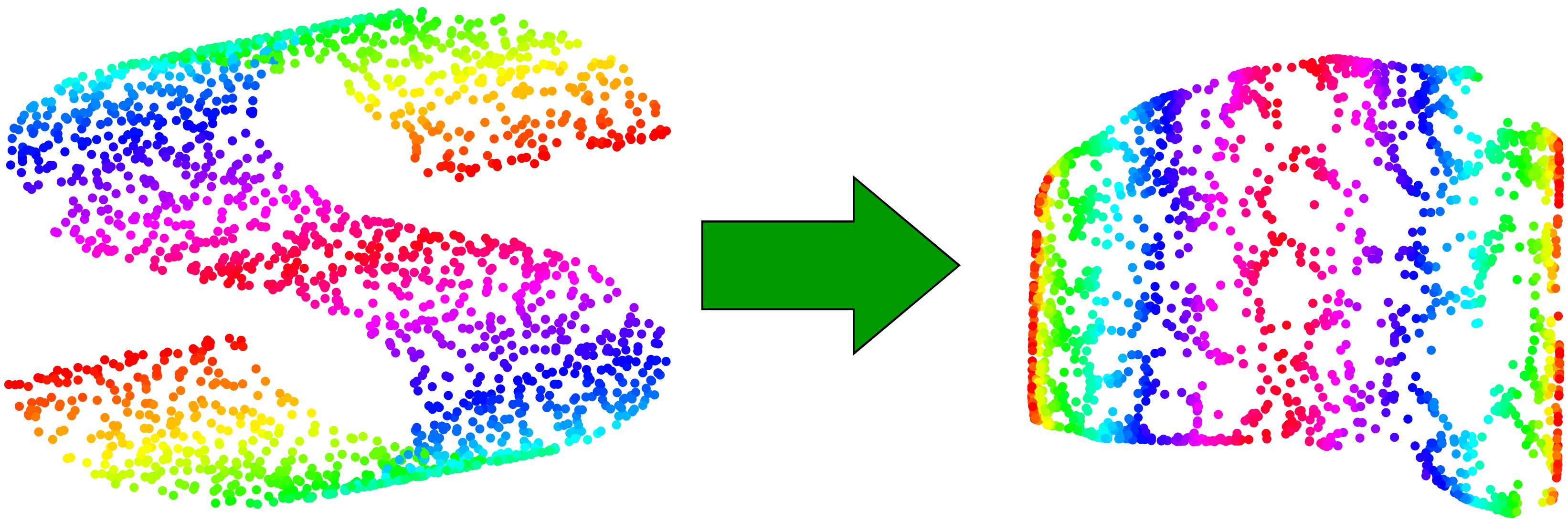


近傍数0の点が多い

次元削減

- 本質的な情報を保持したまま次元を減らしたい！

もとの高次元データ　　　　次元削減後のデータ(2d)



- 1～3次元に減らせば，データを可視化できる.
- 次元削減の基本的な仮定:
  - 手持ちの高次元データはある意味で冗長である

線形次元削減

- （高次元）標本:

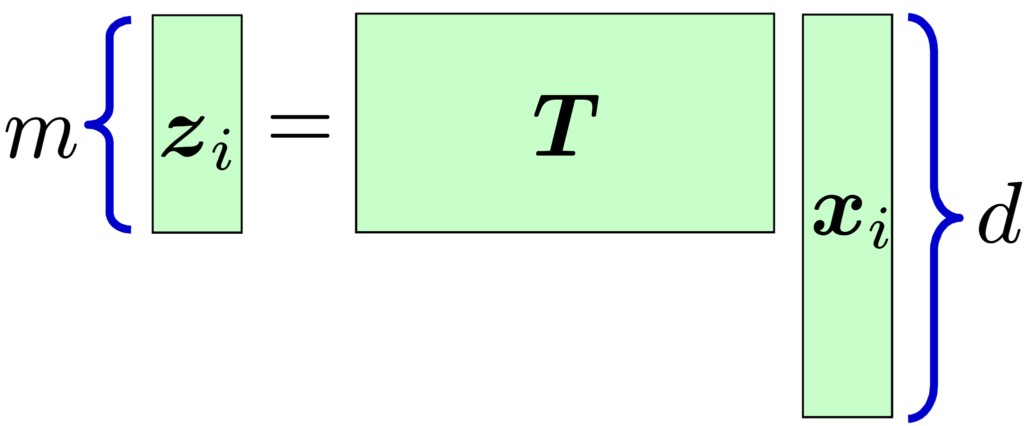
$$\{\boldsymbol{x}_i\}_{i=1}^{n}, \quad \boldsymbol{x}_i \in \mathbb{R}^d, \quad d \gg 1$$

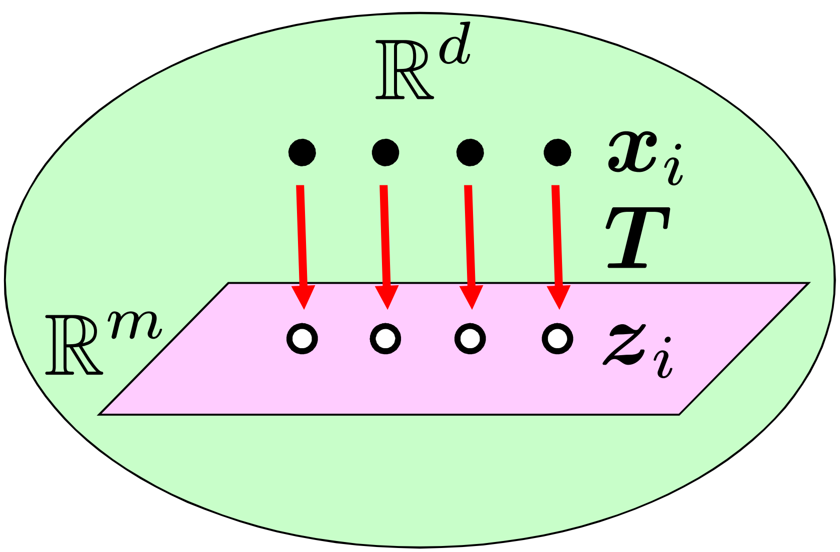
- 埋め込み行列:

$$\boldsymbol{T} \in \mathbb{R}^{m \times d}, \quad 1 \leq m \ll d$$

- 埋め込まれた標本

$$\{\boldsymbol{z}_i\}_{i=1}^{n}, \quad \boldsymbol{z}_i = \boldsymbol{T}\boldsymbol{x}_i \in \mathbb{R}^m$$

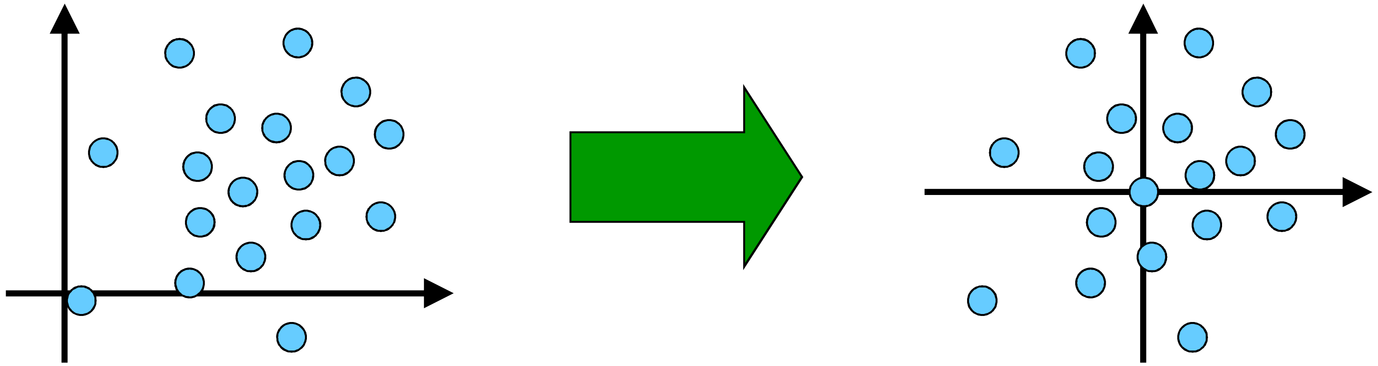




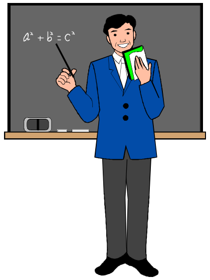
標本の中心化

- 以後，議論を簡単にするため標本を中心化して平均がゼロになるようにしておく

$$\boldsymbol{x}_i \longleftarrow \boldsymbol{x}_i - \frac{1}{n}\sum_{i'=1}^{n}\boldsymbol{x}_{i'}$$



- 分散が二乗平均で表せる



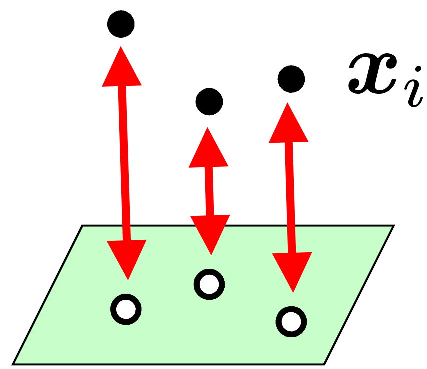
講義の流れ

1. 主成分分析（13章）
2. 局所性保存射影（13章）

主成分分析 (PCA: Principal Component Analysis)

- 考え方：データを表現するうえで最も重要な次元（部分空間）を取り出す

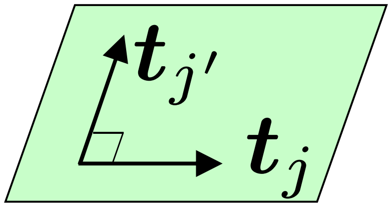
低次元の部分空間に正射影したときにできるだけデータが変化しないようにする



正射影

- $\{\boldsymbol{t}_j \mid \boldsymbol{t}_j \in \mathbb{R}^d\}_{j=1}^m$ ：$m \le d$ 次元部分空間の正規直交基底

$$\boldsymbol{t}_j^\top \boldsymbol{t}_{j'} = \begin{cases} 1 & (j = j') \\ 0 & (j \neq j') \end{cases}$$

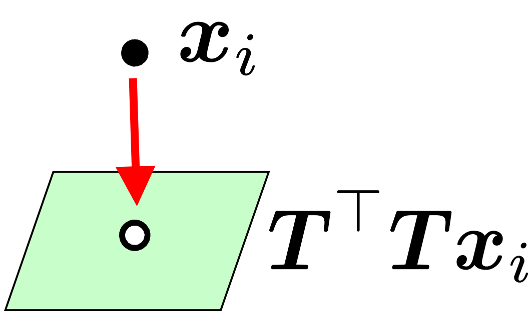


- 行列で表現すると, $\boldsymbol{T}\boldsymbol{T}^\top = \boldsymbol{I}_m$

$$\boldsymbol{T} = (\boldsymbol{t}_1, \ldots, \boldsymbol{t}_m)^\top \in \mathbb{R}^{m \times d}$$

- 標本 $\boldsymbol{x}_i$ の正射影は

$$\sum_{j=1}^m (\boldsymbol{t}_j^\top \boldsymbol{x}_i)\boldsymbol{t}_j \quad \left(= \boldsymbol{T}^\top \boldsymbol{T} \boldsymbol{x}_i\right)$$



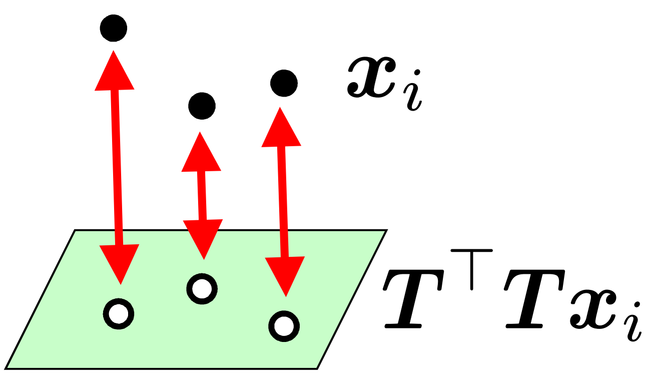
主成分分析の規準

- 射影誤差の和を最小にする:

$$\boldsymbol{T}_{\mathrm{PCA}} = \mathop{\mathrm{argmin}}_{\boldsymbol{T}\in\mathbb{R}^{m\times d}} \left[\sum_{i=1}^{n} \|\boldsymbol{T}^\top \boldsymbol{T}\boldsymbol{x}_i - \boldsymbol{x}_i\|^2 \right]$$

$$\text{subject to } \boldsymbol{T}\boldsymbol{T}^\top = \boldsymbol{I}_m$$

（正規直交性の拘束条件をつける）



数学演習

- $\boldsymbol{T}\boldsymbol{T}^\top = \boldsymbol{I}_m$ および $\displaystyle\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{x}_i = 0$ に対して

  次式を証明せよ

$$\sum_{i=1}^{n}\|\boldsymbol{T}^\top\boldsymbol{T}\boldsymbol{x}_i - \boldsymbol{x}_i\|^2 = -\mathrm{tr}\left(\boldsymbol{T}\boldsymbol{C}\boldsymbol{T}^\top\right) + \mathrm{tr}\left(\boldsymbol{C}\right)$$

- $\mathrm{tr}\left(\cdot\right)$: 行列のトレース
- $\boldsymbol{C} = \displaystyle\sum_{i=1}^{n}\boldsymbol{x}_i\boldsymbol{x}_i^\top$

  標本の散布行列
  （正規化されていない共分散行列）
- ヒント： $\boldsymbol{a}^\top\boldsymbol{b} = \mathrm{tr}\left(\boldsymbol{b}\boldsymbol{a}^\top\right)$

解答例

$$\sum_{i=1}^{n} \|\boldsymbol{T}^\top \boldsymbol{T}\boldsymbol{x}_i - \boldsymbol{x}_i\|^2$$

$$= \sum_{i=1}^{n} \boldsymbol{x}_i^\top \boldsymbol{T}^\top \boldsymbol{T}\boldsymbol{T}^\top \boldsymbol{T}\boldsymbol{x}_i - 2\sum_{i=1}^{n} \boldsymbol{x}_i^\top \boldsymbol{T}^\top \boldsymbol{T}\boldsymbol{x}_i + \sum_{i=1}^{n} \boldsymbol{x}_i^\top \boldsymbol{x}_i$$

$\boldsymbol{T}^\top \boldsymbol{T}\boldsymbol{T}^\top \boldsymbol{T} = \boldsymbol{T}^\top \boldsymbol{T}$: 2回射影しても変わらない

$$= -\sum_{i=1}^{n} \boldsymbol{x}_i^\top \boldsymbol{T}^\top \boldsymbol{T}\boldsymbol{x}_i + \sum_{i=1}^{n} \boldsymbol{x}_i^\top \boldsymbol{x}_i$$

$$\boldsymbol{a}^\top \boldsymbol{b} = \mathrm{tr}\left(\boldsymbol{b}\boldsymbol{a}^\top\right)$$

$$= -\sum_{i=1}^{n} \mathrm{tr}\left(\boldsymbol{T}\boldsymbol{x}_i\boldsymbol{x}_i^\top \boldsymbol{T}^\top\right) + \sum_{i=1}^{n} \mathrm{tr}\left(\boldsymbol{x}_i\boldsymbol{x}_i^\top\right)$$

$$= -\mathrm{tr}\left(\boldsymbol{T}\boldsymbol{C}\boldsymbol{T}^\top\right) + \mathrm{tr}\left(\boldsymbol{C}\right)$$

$$\boldsymbol{C} = \sum_{i=1}^{n} \boldsymbol{x}_i\boldsymbol{x}_i^\top$$

固有値分解

- 対称行列 C は必ず以下の形に分解可能

$$C = \sum_{j=1}^{d} \lambda_j \boldsymbol{\xi}_j \boldsymbol{\xi}_j^\top$$

- $\lambda_j, \boldsymbol{\xi}_j$ は固有値，固有ベクトルとよばれ，次の固有方程式の解として与えられる

$$C\boldsymbol{\xi} = \lambda \boldsymbol{\xi}$$

- 固有ベクトル $\boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_d$ は互いに直交する：

$$\boldsymbol{\xi}_j^\top \boldsymbol{\xi}_{j'} = 0 \qquad \text{for } j \neq j'$$

  - 以降，規格化済みと仮定：$\|\boldsymbol{\xi}_j\| = 1$
- Octaveではeig関数で計算できる

固有値分解

$$\boldsymbol{C} = \sum_{j=1}^{d} \lambda_j \boldsymbol{\xi}_j \boldsymbol{\xi}_j^\top$$

- 固有値分解を用いれば，逆行列は次式で，

$$\boldsymbol{C}^{-1} = \sum_{j=1}^{d} \lambda_j^{-1} \boldsymbol{\xi}_j \boldsymbol{\xi}_j^\top \qquad \forall \lambda_j \neq 0$$

行列の平方根（の1つ）は次式で求められる

$$\boldsymbol{C}^{1/2} = \sum_{j=1}^{d} \lambda_j^{1/2} \boldsymbol{\xi}_j \boldsymbol{\xi}_j^\top \qquad \forall \lambda_j > 0$$

- Octaveではinv関数，sqrtm関数で計算できる

主成分分析の解

$$\boldsymbol{T}_{\mathrm{PCA}} = \operatorname*{argmax}_{\boldsymbol{T}\in\mathbb{R}^{m\times d}} \mathrm{tr}\left(\boldsymbol{T}\boldsymbol{C}\boldsymbol{T}^\top\right)$$

$$\text{subject to } \boldsymbol{T}\boldsymbol{T}^\top = \boldsymbol{I}_m$$

- 主成分分析の解は次式で与えられる:

$$\boldsymbol{T}_{\mathrm{PCA}} = \boldsymbol{U}(\boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_m)^\top$$

  - $\boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_m$:固有値問題 $\boldsymbol{C}\boldsymbol{\xi} = \lambda\boldsymbol{\xi}$ の固有値 $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_d$ に対応する固有ベクトル
  - $\boldsymbol{U}$:任意の $m \times m$ 直交行列 $\boldsymbol{U}^{-1} = \boldsymbol{U}^\top$
- 通常は $\boldsymbol{U} = \boldsymbol{I}_m$ を用いる

証明

- ラグランジュ関数:

$$L(\boldsymbol{T}, \boldsymbol{\Delta}) = \mathrm{tr}(\boldsymbol{T}\boldsymbol{C}\boldsymbol{T}^\top) - \mathrm{tr}((\boldsymbol{T}\boldsymbol{T}^\top - \boldsymbol{I}_m)\boldsymbol{\Delta})$$

$\boldsymbol{\Delta}$:ラグランジュ未定乗数(対称行列)

- 最適性条件:
  - $\dfrac{\partial L}{\partial \boldsymbol{T}} = 2\boldsymbol{T}\boldsymbol{C} - 2\boldsymbol{\Delta}\boldsymbol{T} = 0$ より $\boldsymbol{C}\boldsymbol{T}^\top = \boldsymbol{T}^\top\boldsymbol{\Delta}$ (1)
  - 元の制約より $\boldsymbol{T}\boldsymbol{T}^\top = \boldsymbol{I}_m$ (2)

- $\boldsymbol{\Delta}$ の固有値分解: $\boldsymbol{\Delta} = \boldsymbol{U}\boldsymbol{\Gamma}\boldsymbol{U}^\top$ (3)

$\boldsymbol{U}$:直交行列
$\boldsymbol{\Gamma}$:対角行列

- (1) と (3) より $\boldsymbol{C}\boldsymbol{T}^\top = \boldsymbol{T}^\top\boldsymbol{U}\boldsymbol{\Gamma}\boldsymbol{U}^\top$ (4)
- (4) の右から $\boldsymbol{U}$ をかけると

$$\boldsymbol{C}\boldsymbol{T}^\top\boldsymbol{U} = \boldsymbol{T}^\top\boldsymbol{U}\boldsymbol{\Gamma} \quad (5)$$

証明(続き)

$$\boldsymbol{C}\boldsymbol{T}^\top\boldsymbol{U} = \boldsymbol{T}^\top\boldsymbol{U}\boldsymbol{\Gamma} \quad (5)$$

- $\boldsymbol{T}^\top\boldsymbol{U} = (\boldsymbol{\xi}'_1, \ldots, \boldsymbol{\xi}'_m)$ とおいて(5)の各列を比べると
  - $\boldsymbol{C}\boldsymbol{\xi}'_i = \gamma_i\boldsymbol{\xi}'_i$ ($\boldsymbol{\Gamma} = \mathrm{diag}\left(\gamma_1, \ldots, \gamma_m\right)$ とおいた)
  - 固有方程式より各 ξ'_i は C の固有ベクトル

- $(\boldsymbol{T}^\top\boldsymbol{U})^\top(\boldsymbol{T}^\top\boldsymbol{U}) = \boldsymbol{U}^\top\boldsymbol{T}\boldsymbol{T}^\top\boldsymbol{U} = \boldsymbol{I}_m$ より各 ξ'_i は相異なる固有ベクトル: $\boldsymbol{\xi}'_i = \boldsymbol{\xi}_{k_i}$　$\forall_{i\neq i'} k_i \neq k_{i'}$

$$\boldsymbol{\Gamma} = \mathrm{diag}\left(\lambda_{k_m}, \ldots, \lambda_{k_m}\right) \quad (6)$$

証明(続き)

$$\boldsymbol{T}\boldsymbol{T}^\top = \boldsymbol{I}_m \quad (2)$$

$$\boldsymbol{C}\boldsymbol{T}^\top = \boldsymbol{T}^\top \boldsymbol{U}\boldsymbol{\Gamma}\boldsymbol{U}^\top \quad (4)$$

$$\boldsymbol{\Gamma} = \mathrm{diag}\left(\lambda_{k_m}, \ldots, \lambda_{k_m}\right) \quad (6) \qquad \forall_{i \neq i'} k_i \neq k_{i'}$$

- 以上をまとめると

$$\mathrm{tr}\left(\boldsymbol{T}\boldsymbol{C}\boldsymbol{T}^\top\right) = \mathrm{tr}\left(\boldsymbol{T}\boldsymbol{T}^\top \boldsymbol{U}\boldsymbol{\Gamma}\boldsymbol{U}^\top\right) = \mathrm{tr}\left(\boldsymbol{U}\boldsymbol{\Gamma}\boldsymbol{U}^\top\right) = \mathrm{tr}\left(\boldsymbol{\Gamma}\right) = \sum_{i=1}^{m} \lambda_{k_i}$$

- $\lambda_1 \geq \cdots \geq \lambda_d$ より $\mathrm{tr}(\boldsymbol{T}\boldsymbol{C}\boldsymbol{T}^\top)$ は $k_i = i$ で最大
- よって

$$\boldsymbol{T}_{\mathrm{PCA}} = \boldsymbol{U}(\boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_m)^\top$$

主成分分析の解の求め方

1. 固有値問題を解く:

$$\boldsymbol{C}\boldsymbol{\xi} = \lambda\boldsymbol{\xi}$$

標本の散布行列

$$\boldsymbol{C} = \sum_{i=1}^{n} \boldsymbol{x}_i \boldsymbol{x}_i^\top$$

- 固有値を降順にソート: $\lambda_1 \geq \cdots \geq \lambda_d$
- 固有ベクトルを正規化: $\|\boldsymbol{\xi}_j\| = 1$

2. 上位 m 個の固有ベクトルを並べる:

$$\boldsymbol{T}_{\mathrm{PCA}} = \boldsymbol{U}(\boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_m)^\top$$

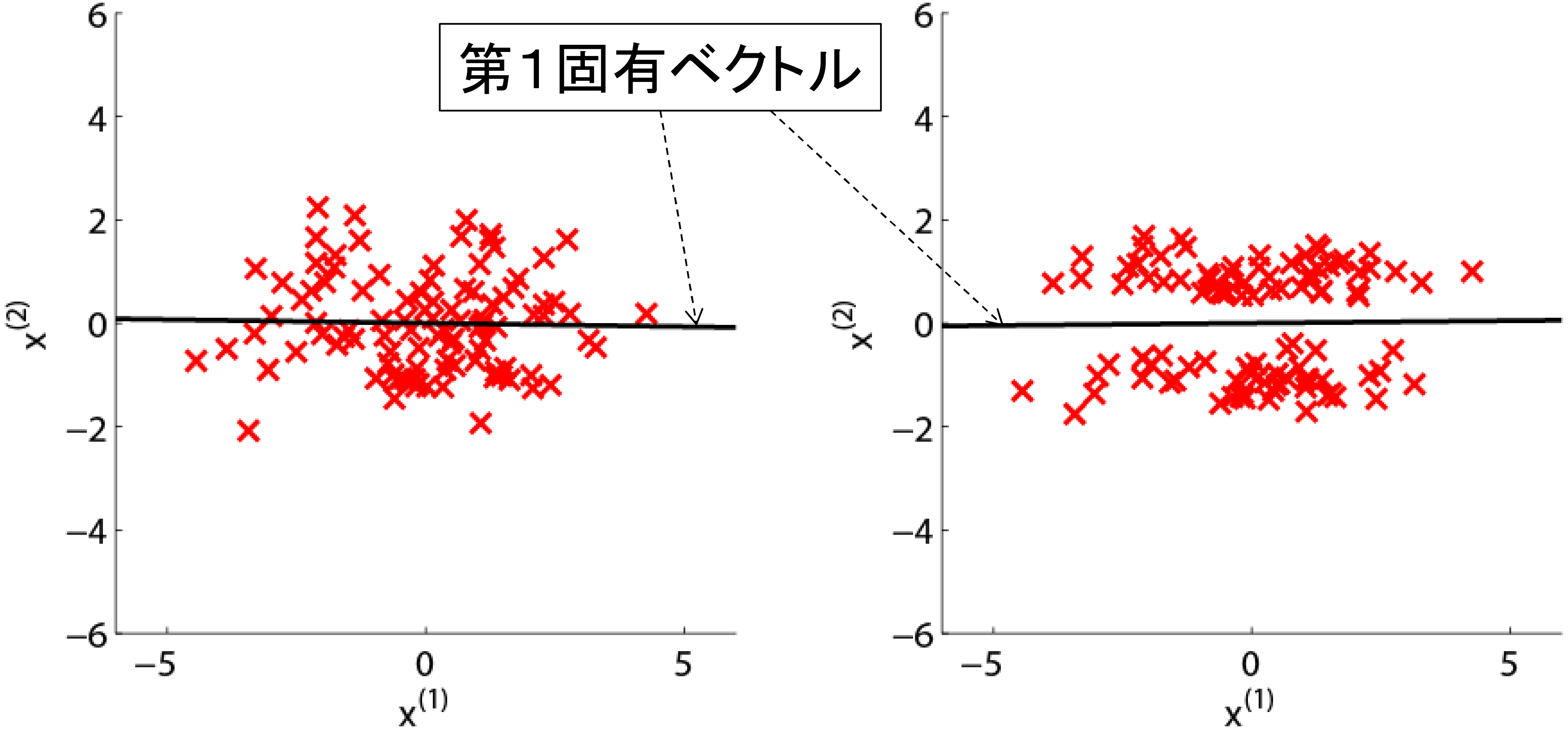
$\boldsymbol{U}$:任意の直交行列

実行例


```
clear all; rand('state',0); randn('state',4);
n=100;
%x=[2*randn(n,1) randn(n,1)];
x=[2*randn(n,1) 2*round(rand(n,1))-1+randn(n,1)/3];
x=x-repmat(mean(x),[n,1]);
[t,v]=eigs(x'*x,1);

figure(1); clf; hold on; axis([-6 6 -6 6])
plot(x(:,1),x(:,2),'rx')
plot(9*[-t(1) t(1)],9*[-t(2) t(2)]);
```

結果の例



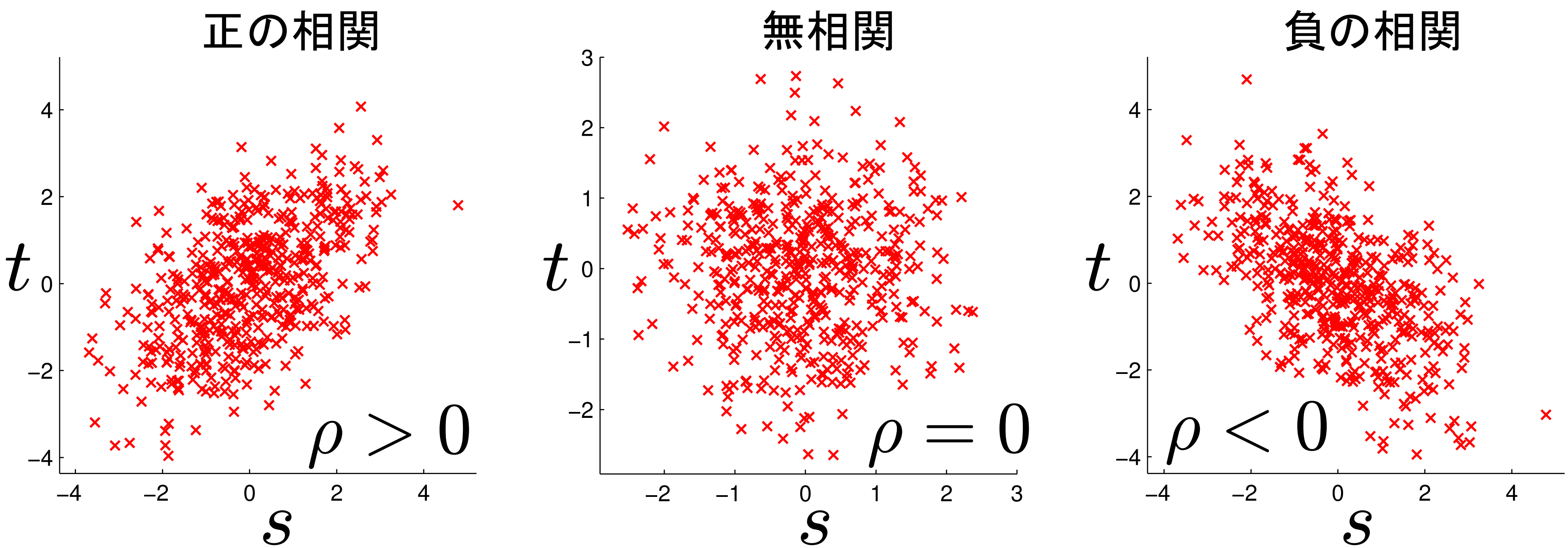
- 主成分分析によって，データの大局的な分布をよく表す部分空間が得られている
- しかし，クラスタ構造のようなデータの局所的な構造は失われることがある

ピアソンの相関係数

- 対データ $\{(s_i, t_i)\}_{i=1}^n$ の相関係数:

$$\rho = \frac{\sum_{i=1}^n (s_i - \overline{s})(t_i - \overline{t})}{\sqrt{\left(\sum_{i=1}^n (s_i - \overline{s})^2\right)\left(\sum_{i=1}^n (t_i - \overline{t})^2\right)}}$$

$$\overline{s} = \frac{1}{n}\sum_{i=1}^n s_i \qquad \overline{t} = \frac{1}{n}\sum_{i=1}^n t_i$$



数学演習

- 主成分分析によって得られた次元削減後の標本

$$\boldsymbol{z}_i = \boldsymbol{T}_{\mathrm{PCA}} \boldsymbol{x}_i \qquad i = 1, \ldots, n \qquad \boldsymbol{C} = \sum_{i=1}^{n} \boldsymbol{x}_i \boldsymbol{x}_i^\top$$

$$\boldsymbol{T}_{\mathrm{PCA}} = (\boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_m)^\top \qquad \boldsymbol{C}\boldsymbol{\xi} = \lambda\boldsymbol{\xi}$$

の散布行列が次式で与えられることを証明せよ

$$\sum_{i=1}^{n} \boldsymbol{z}_i \boldsymbol{z}_i^\top = \mathrm{diag}(\lambda_1, \ldots, \lambda_m) \qquad \lambda_1 \geq \cdots \geq \lambda_d$$

- これより，　$\boldsymbol{z} = (z^{(1)}, \ldots, z^{(m)})^\top$ の各要素は無相関であることがわかる

解答例

$$\sum_{i=1}^{n} \boldsymbol{z}_i \boldsymbol{z}_i^\top = \boldsymbol{T}_{\mathrm{PCA}} \left(\sum_{i=1}^{n} \boldsymbol{x}_i \boldsymbol{x}_i^\top \right) \boldsymbol{T}_{\mathrm{PCA}}^\top$$

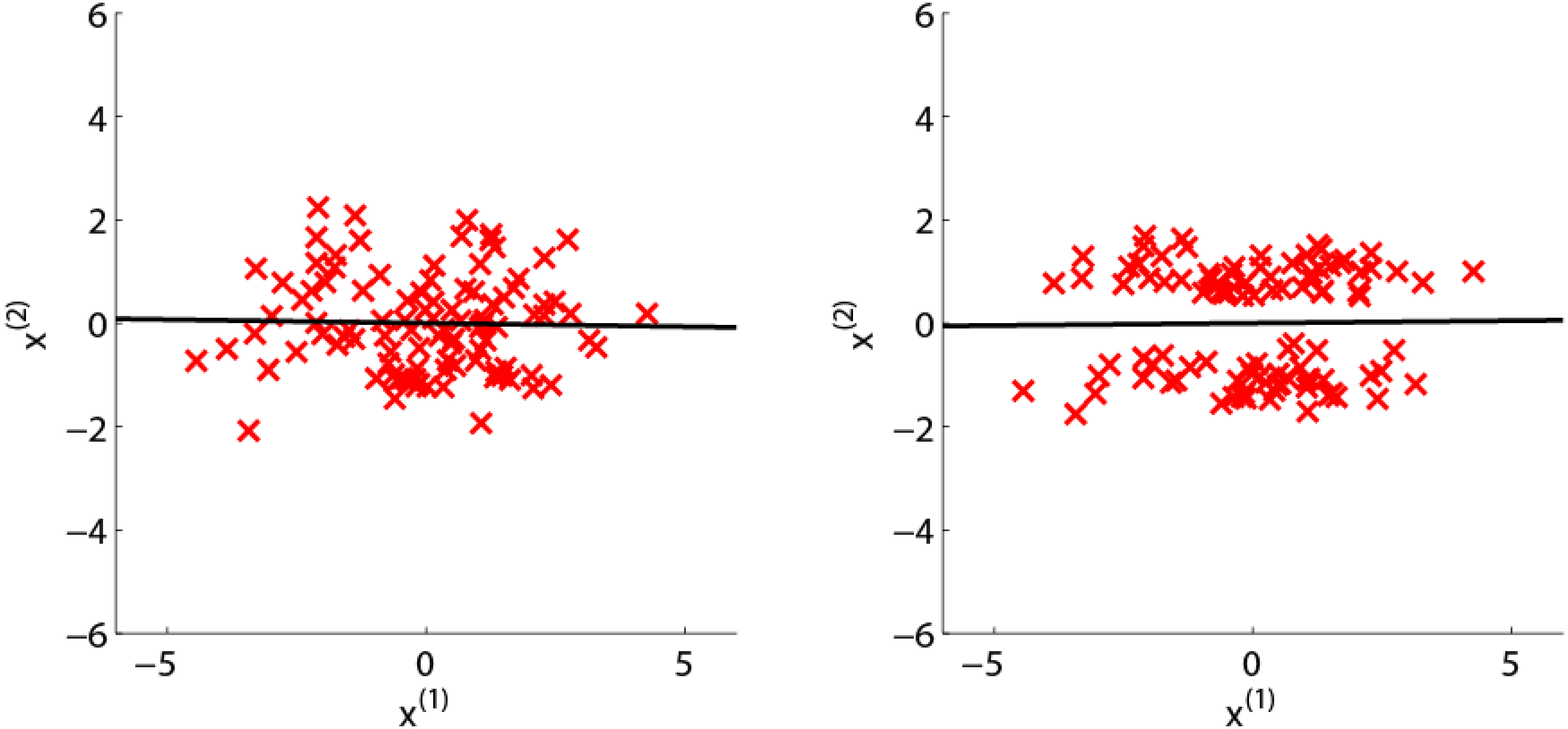
$$= \boldsymbol{T}_{\mathrm{PCA}} \boldsymbol{C} \boldsymbol{T}_{\mathrm{PCA}}^\top$$

$$= \boldsymbol{T}_{\mathrm{PCA}} \left(\sum_{j=1}^{d} \lambda_i \boldsymbol{\xi}_i \boldsymbol{\xi}_i^\top \right) \boldsymbol{T}_{\mathrm{PCA}}^\top$$

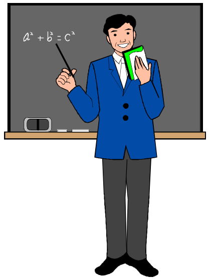
$$\boldsymbol{T}_{\mathrm{PCA}} = (\boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_m)^\top$$

$$= \mathrm{diag}\left(\lambda_1, \ldots, \lambda_m\right)$$

主成分分析:まとめ



- 固有ベクトルから解析的に解が求まる
- データの大局的な分布を捉えられる
- データは無相関化される
- クラスタ構造は失われることがある



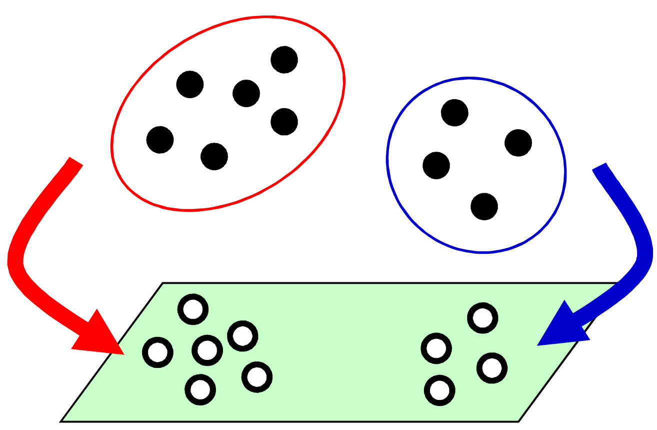
講義の流れ

1. 主成分分析(13章)
2. 局所性保存射影(13章)

局所性保存射影 (LPP: Locality Preserving Projection)

- 考え方:データの局所的な構造をできるだけ保存したい

もとの空間で近くにある標本を近くに埋め込む



類似度行列

■ 類似度行列 $\boldsymbol{W}$:

- $\boldsymbol{x}_i$ と $\boldsymbol{x}_{i'}$ が近い: $W_{i,i'} \to 1$
- $\boldsymbol{x}_i$ と $\boldsymbol{x}_{i'}$ が遠い: $W_{i,i'} \to 0$

$W_{i,i'} = W_{i',i}$

$^{\exists}\boldsymbol{W}^{-1}$

■ 例:

- 距離ベース: $$W_{i,i'} = \exp\left(-\frac{\|\boldsymbol{x}_i - \boldsymbol{x}_{i'}\|^2}{2t^2}\right)$$

$t > 0$

$k \in \mathbb{N}$

- 近傍ベース: $$W_{i,i'} = \begin{cases} 1 & \text{if } \boldsymbol{x}_i \in kNN(\boldsymbol{x}_{i'}) \\ & \quad\quad \text{or } \boldsymbol{x}_{i'} \in kNN(\boldsymbol{x}_i) \\ 0 & \text{otherwise} \end{cases}$$

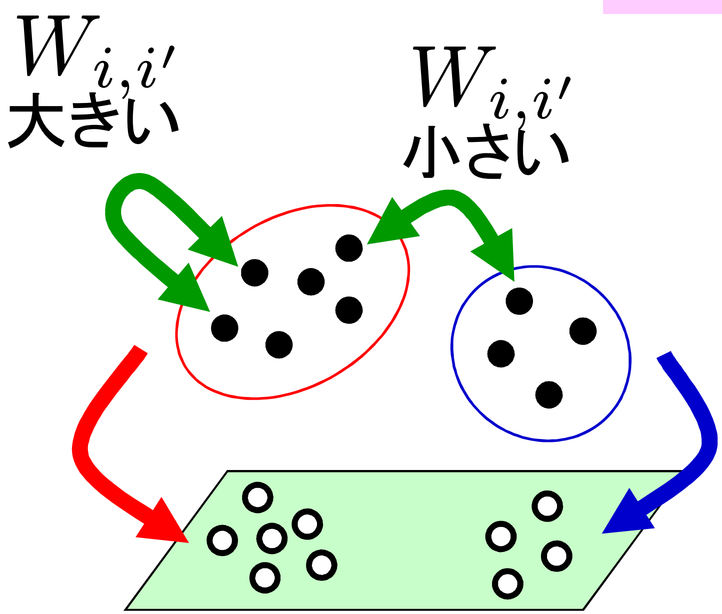
局所性保存射影の規準

■ もとの空間で近くにある標本を近くに埋め込む:

$$\boldsymbol{T}_{\mathrm{LPP}} = \underset{\boldsymbol{T}\in\mathbb{R}^{m\times d}}{\mathrm{argmin}} \sum_{i,i'=1}^{n} W_{i,i'} \|\boldsymbol{T}\boldsymbol{x}_i - \boldsymbol{T}\boldsymbol{x}_{i'}\|^2$$

$$\text{subject to } \boldsymbol{T}\boldsymbol{X}\boldsymbol{D}\boldsymbol{X}^\top\boldsymbol{T}^\top = \boldsymbol{I}_m$$

解の縮退を防ぐための
拘束条件(後述)



$$\boldsymbol{X} = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n)$$

$$\boldsymbol{D} = \mathrm{diag}\left(\sum_{i'=1}^{n} W_{i,i'}\right)$$

拘束条件の直感的な意味

- 拘束条件: $\boldsymbol{T}\boldsymbol{X}\boldsymbol{D}\boldsymbol{X}^\top\boldsymbol{T}^\top = \boldsymbol{I}_m$ $\quad \boldsymbol{D} = \mathrm{diag}\left(\sum_{i'=1}^{n} W_{i,i'}\right)$
- $\boldsymbol{T} = \boldsymbol{0}$ なら目的関数をゼロにできる
- $\boldsymbol{T}\boldsymbol{X} = \boldsymbol{T}(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n)$ を「膨らませて」拘束条件を満たすようにする
- D_{ii} が大きいとき, $\boldsymbol{T}\boldsymbol{x}_i$を大きくすると一気に $\boldsymbol{T}\boldsymbol{X}\boldsymbol{D}\boldsymbol{X}^\top\boldsymbol{T}^\top$ を大きくできる
- D_{ii} は点 $\boldsymbol{x}_i$に近い点の多さ≒重要度に対応
  - 重要な点の$\boldsymbol{T}\boldsymbol{x}_i$ が強調されるような$\boldsymbol{T}$が最適解として表れやすい

式変形

- （前回演習とほぼ同様に）次式が成り立つ

$$\sum_{i,i'=1}^{n} W_{i,i'} \|\boldsymbol{T}\boldsymbol{x}_i - \boldsymbol{T}\boldsymbol{x}_{i'}\|^2 = 2\mathrm{tr}\left(\boldsymbol{T}\boldsymbol{X}\boldsymbol{L}\boldsymbol{X}^\top\boldsymbol{T}^\top\right)$$

証明は宿題

$$\boldsymbol{X} = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n)$$

$$\boldsymbol{L} = \boldsymbol{D} - \boldsymbol{W}$$

$$\boldsymbol{D} = \mathrm{diag}\left(\sum_{i'=1}^{n} W_{i,i'}\right)$$

一般化固有値問題

$$\boldsymbol{A}\boldsymbol{\xi} = \lambda \boldsymbol{B}\boldsymbol{\xi}$$

$\boldsymbol{B}$:正値対称行列

正値＝全ての固有値が正

- $\boldsymbol{\phi} = \boldsymbol{B}^{1/2}\boldsymbol{\xi} \Leftrightarrow \boldsymbol{B}^{-1/2}\boldsymbol{\phi} = \boldsymbol{\xi}$ とおいて左から$\boldsymbol{B}^{-1/2}$をかければ通常の固有値問題に還元できる

$$\boldsymbol{B}^{-1/2}\boldsymbol{A}\boldsymbol{B}^{-1/2}\boldsymbol{\phi} = \lambda\boldsymbol{\phi}$$

- 固有ベクトルの直交性

$$\boldsymbol{\phi}_j^\top \boldsymbol{\phi}_{j'} = 0 \quad \text{for } j \neq j'$$

より，一般化固有ベクトルは$\boldsymbol{B}$-直交性を満たす

$$\boldsymbol{\xi}_j^\top \boldsymbol{B}\boldsymbol{\xi}_{j'} = 0 \quad \text{for } j \neq j'$$

局所性保存射影の解の求め方

$$\boldsymbol{T}_{\mathrm{LPP}} = \mathop{\mathrm{argmin}}_{\boldsymbol{T}\in\mathbb{R}^{m\times d}} \mathrm{tr}\left(\boldsymbol{T}\boldsymbol{X}\boldsymbol{L}\boldsymbol{X}^\top\boldsymbol{T}^\top\right)$$

$$\text{subject to } \boldsymbol{T}\boldsymbol{X}\boldsymbol{D}\boldsymbol{X}^\top\boldsymbol{T}^\top = \boldsymbol{I}_m$$

1. 一般化固有値問題を解く:

$$\boldsymbol{L} = \boldsymbol{D} - \boldsymbol{W}$$

$$\boldsymbol{X}\boldsymbol{L}\boldsymbol{X}^\top\boldsymbol{\xi} = \lambda\boldsymbol{X}\boldsymbol{D}\boldsymbol{X}^\top\boldsymbol{\xi}$$

$$\boldsymbol{D} = \mathrm{diag}\left(\sum_{i'=1}^{n} W_{i,i'}\right)$$

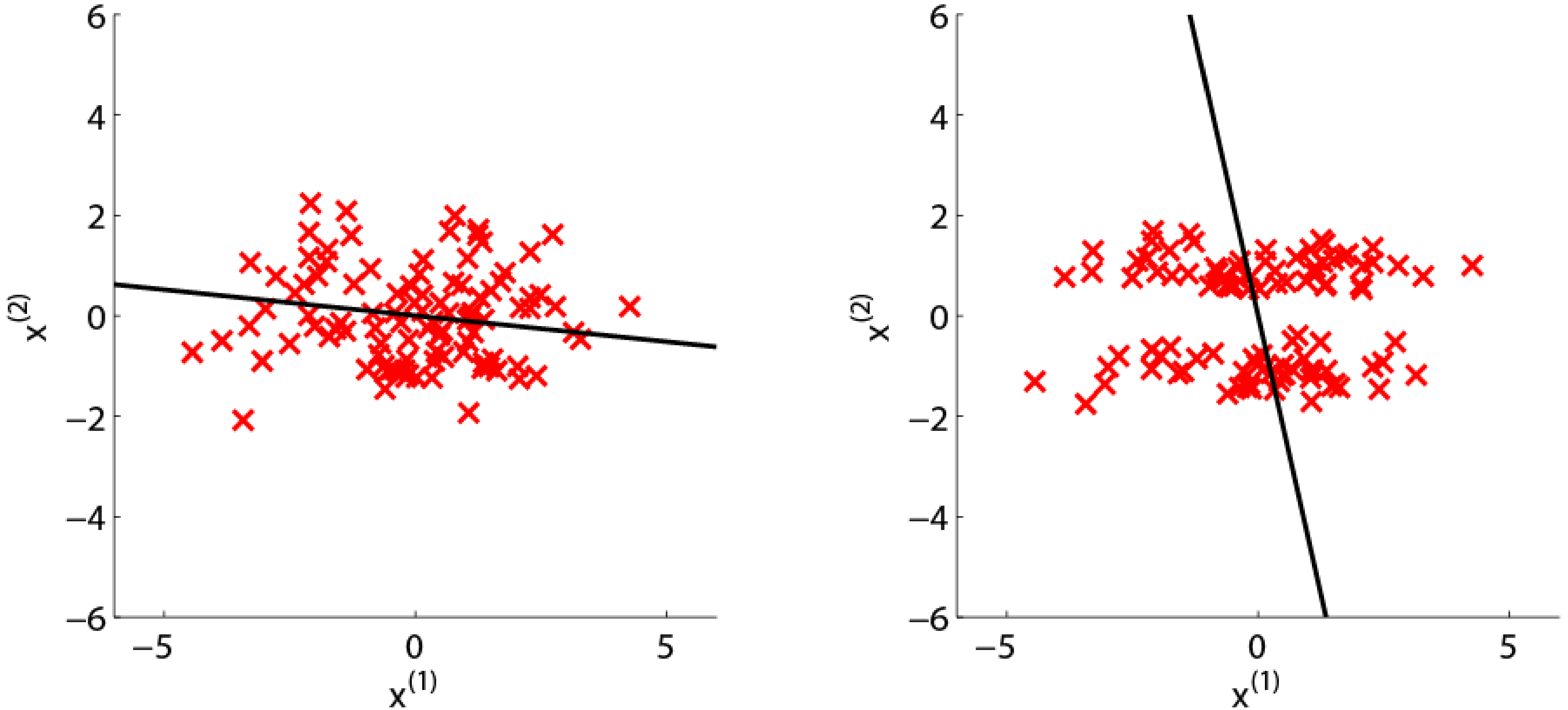
$$\lambda_1 \geq \cdots \geq \lambda_d$$

$$\boldsymbol{\xi}_j^\top\boldsymbol{X}\boldsymbol{D}\boldsymbol{X}^\top\boldsymbol{\xi}_j = 1$$

2. 下位 m 個の一般化固有ベクトルを並べる:

$$\boldsymbol{T}_{\mathrm{LPP}} = (\boldsymbol{\xi}_d, \boldsymbol{\xi}_{d-1}, \ldots, \boldsymbol{\xi}_{d-m+1})^\top$$

実行例

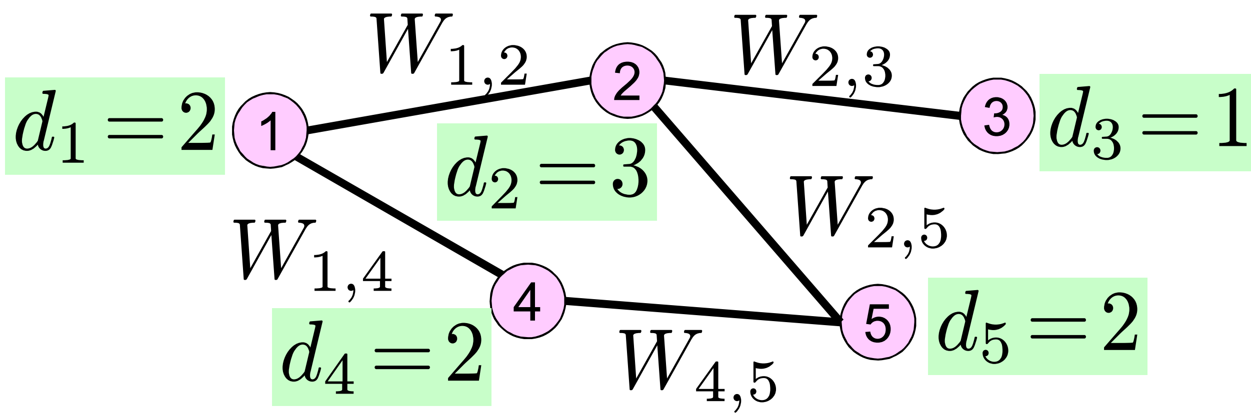


- 局所性保存射影によって，クラスタ構造をうまく保存できている

実装は宿題

グラフ理論

- **グラフ**: 頂点と枝からなる
- **隣接行列** $\boldsymbol{W}$: $W_{i,i'}$ は頂点 i, i' 間の枝の数
- **頂点次数** d_i: 頂点 i につながっている枝の数



- **グラフラプラス行列**:

$$L_{i,i'} = \begin{cases} d_i & (i = i') \\ -1 & (i \neq i' \text{ and } W_{i,i'} > 0) \\ 0 & (\text{otherwise}) \end{cases}$$

グラフ理論

- 近傍に基づく類似度行列を考える

$$W_{i,i'} = \begin{cases} 1 & \text{if } \boldsymbol{x}_i \in kNN(\boldsymbol{x}_{i'}) \\ & \quad\quad \text{or } \boldsymbol{x}_{i'} \in kNN(\boldsymbol{x}_i) \\ 0 & \text{otherwise} \end{cases}$$

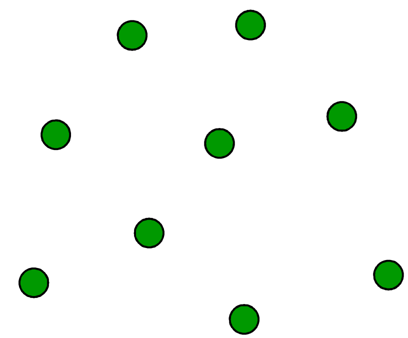
- これに対応するグラフを考えると
  - $\boldsymbol{W}$: 隣接行列
  - $\boldsymbol{D}$: 頂点次数からなる対角行列
  - $\boldsymbol{L}$: グラフラプラス行列

$$\boldsymbol{D} = \mathrm{diag}\left(\sum_{i'=1}^{n} W_{i,i'}\right)$$

$$\boldsymbol{L} = \boldsymbol{D} - \boldsymbol{W}$$

類似度行列の決め方

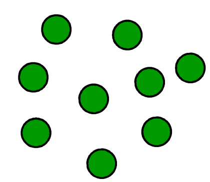
- 埋込結果が**類似度行列の決め方**に依存する.
- **局所スケーリングのヒューリスティクス:**



$$W_{i,i'} = \exp\left(-\frac{\|\boldsymbol{x}_i - \boldsymbol{x}_{i'}\|^2}{t_i t_{i'}}\right)$$

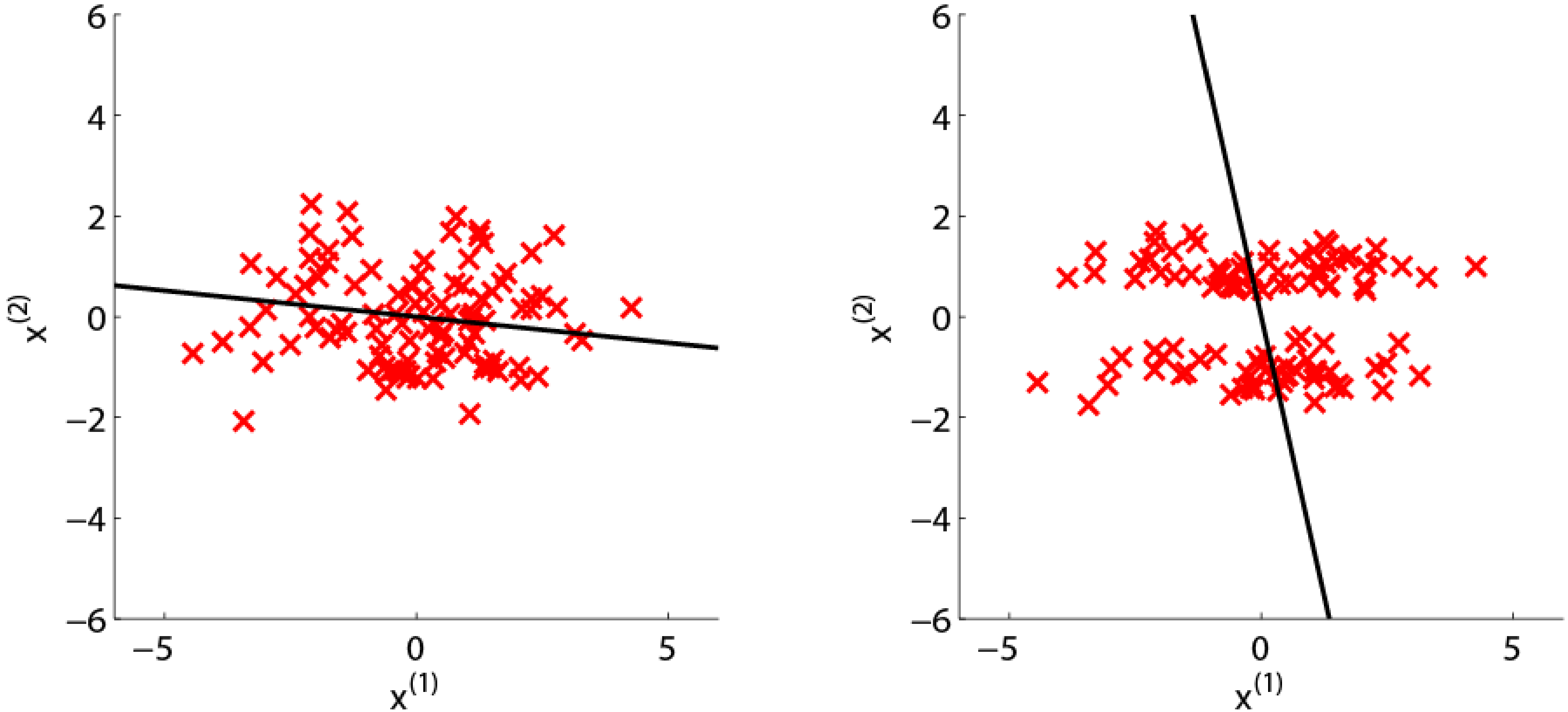
$$t_i = \|\boldsymbol{x}_i - \boldsymbol{x}_i^{(k)}\|$$

$\boldsymbol{x}_i^{(k)}$: $\boldsymbol{x}_i$ の k 番目の近傍の標本

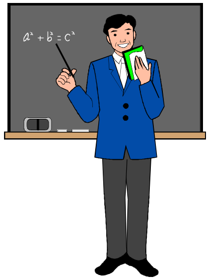


- k の決め方が自明でない

局所性保存射影:まとめ



- 一般化固有ベクトルから解析的に解が求まる
- データの局所的な分布を捉えられる
- 結果が類似度行列の選び方に依存する

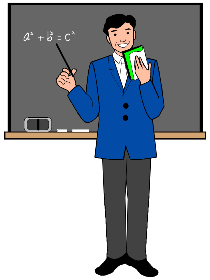


講義の流れ

1. 主成分分析（13章）
2. 局所性保存射影（13章）

まとめ

- 次元の呪い: 入力の次元が増えると学習問題は指数的に難しくなる
- 次元削減: 本質的な情報を保持したままデータの次元を減らす
  - 主成分分析: 元のデータをできるだけ保存
  - 局所性保存射影: クラスタ構造をできるだけ保存



次回の予告

- 教師付き学習における線形次元削減（17章）

宿題1

- 次式を証明せよ

$$\sum_{i,i'=1}^{n} W_{i,i'} \|\boldsymbol{T}\boldsymbol{x}_i - \boldsymbol{T}\boldsymbol{x}_{i'}\|^2 = 2\mathrm{tr}\left(\boldsymbol{T}\boldsymbol{X}\boldsymbol{L}\boldsymbol{X}^\top \boldsymbol{T}^\top\right)$$

$$\boldsymbol{X} = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n)$$

$$\boldsymbol{L} = \boldsymbol{D} - \boldsymbol{W}$$

$$\boldsymbol{D} = \mathrm{diag}\left(\sum_{i'=1}^{n} W_{i,i'}\right)$$

宿題2

- 適当な類似度行列に対して局所性保存射影を実装せよ:

```
clear all; rand('state',0); randn('state',4);
n=100;
% x=[2*randn(n,1) randn(n,1)];
x=[2*randn(n,1) 2*round(rand(n,1))-1+randn(n,1)/3];
```